UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/565,570	07/21/2006	Robert W. Morris	30004-004US1	2103

69713 7590 06/29/2010

OCCHIUTI ROHLICEK & TSAO, LLP
10 FAWCETT STREET
CAMBRIDGE, MA 02138

EXAMINER	
BORSETTI, GREG	

ART UNIT	PAPER NUMBER
2626	

NOTIFICATION DATE	DELIVERY MODE
06/29/2010	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

INFO@ORTPATENT.COM

Office Action Summary	Application No.	Applicant(s)	
	10/565,570	MORRIS, ROBERT W.	
	Examiner	Art Unit	
	GREG A. BORSETTI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *16 June 2010*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-19* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-19* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Claims 1-19 are pending.

2. The 35 USC 101 rejection to claim 1 has been withdrawn.

Response to Arguments

3. Applicant's arguments filed 6/16/2010 have been fully considered but they are not persuasive.

4. Applicant argues "The applicant does not agree...performs the "receiving" step of claim 1." (Remarks, Page 9, ¶ 2) The Examiner disagrees. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., ...receiving input from a user identifying at least two portions...the spoken event of interest in the first set of audio signals...) are not recited in the previously rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). These limitations have been added and will be examined on the merits below.

5. Applicant further argues "there is no user- identification in a first set of audio signals...the spoken event of interest in a second audio signal" (Remarks, pages 9-10) The Examiner disagrees. Inputting a query is a user-identification of a query because it is known that if a person is inputting the query that they are actively seeking a result based on their query. The combination of Wolf with Cardillo teaches a speech based

input where a speech input is a user-indentified audio query. Therefore, the claim language is taught. The argument is not persuasive.

6. Applicant further requests that the points on page 10 are specifically addressed. They will be addressed in the rejection below.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7. Claims 1-4, 8-9, 12-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cardillo et al. (NPL Document Phonetic Searching vs. LVCSR: How to Find What You Really Want in Audio Archives") in view of Wolf et al. (US PGPUB #20030204492)

As per claim 1, Cardillo teaches:

processing, by a query recognizer of a word spotting system, each identified portion of the first set of audio signals to generate a corresponding subword unit representation of the identified portion; (Page 12, column 1, ...words or phrases... a phrase would include at least two portions (words), also the query

representations can be phonetic and would thus also teach phonetic portions. Further, *...A phonetic dictionary is probed for each word within the query term...*)

accepting, by a word spotting engine of the word spotting system, data representing the unknown speech in a second audio signal; and (Fig. 1, preprocessing phase and generation of the search track. Page 11, column 2, *...A phonetic grammar likewise depends upon the natural language in use (particularly the set of phonemes used to represent basic sounds and meanings of the input speech). This grammar is used to identify likely end points of words in the input speech...*)

Cardillo fails to fully teach, but Wolf supplements:

receiving input from a user identifying at least two portions of a first set of audio signals as being of interest to the user, wherein the input includes a first input from the user identifying a first instance of a spoken event of interest in the first set of audio signals and a second input from the user identifying a second instance of the spoken event of interest in the first set of audio signals; (Cardillo, Page 12, column 1, *...words or phrases...* a phrase would include at least two portions (words), also the query representations can be phonetic and would thus also teach phonetic portions. Wolf further teaches that the query may be spoken.

The Examiner will now address the specific limitations. 1) A first set of audio signals is taught by Cardillo in the temporal operators where "brain cancer" and "cell phone" are a first set of audio signals. 2) A spoken event of interest is the combination of brain cancer and cell phone because the person presumably looking for how cell

phone may affect brain cancer. It should be noted that Wolf then provides the spoken queries. 3) A first instance of a spoken event of interest in a first set of audio signals is taught by "brain cancer", this is further suggested by Wolf which teaches the spoken queries. 4) A second instance of the spoken event of interest in the first set of audio signals is taught by "cell phone", this is further suggested by Wolf which teaches the spoken queries. 5) A first input from a user identifying (3) is again taught by "brain cancer" for the reasons above, it is an instance of the query for the spoken event of interest. 6) A second input from a user identifying (4) is again taught by "cell phone" for the reasons above, it is an instance of the query for the spoken event of interest. 7) A second audio signal is taught by Cardillo in the search track.)

forming, by the query recognizer of a word spotting system, a representation of the spoken event of interest, wherein the forming includes combining the subword unit representations of the respective identified portions of the first set of audio signals; (Cardillo, Page 15, column 2, teaches multiple word queries which are two or more instances of a event of interest. They are performed in a phonetic search therefore they comprise at least one sequence of phonemes each. Again, Cardillo fails to specifically teach a spoken event of interest. Wolf, ¶ 0054, *...a phoneme lattice, as described above, can also be used for devices with limited resources...In the case where the recognizer is part of the input device, e.g., a cell phone, the lattices can be forwarded to the search engine 190...* Furthemore, ¶ 0022, *...If phonemes are used, then it is possible to handle words that sound the same but have different meaning...* A phoneme lattice (combination) is used to represent multiple pronunciations of words.)

locating, by the word spotting engine of the word spotting system, putative instances of the spoken event of interest in the second audio signal using the representation of the spoken event of interest, wherein the locating includes identifying time locations of the second audio signal at which the spoken event of interest is likely to have occurred based on a comparison of the data representing the unknown speech with the representation of the spoken event of interest. (Cardillo teaches in Fig. 1, a searching phase which locates putative instances of a query where ,Page 12, Temporal_Offset teaches a time location. Cardillo fails to specifically teach a spoken event of interest (spoken query). Wolf, abstract, ...*A spoken query is represented as a lattice*...)

It would have been obvious to someone of ordinary skill in the art to combine Wolf with Cardillo to avoid losing information and adding ambiguity by application of a speech recognition process and performing traditional textual document retrieval. (Wolf, ¶ 0006)

As per claim 2, claim 1 is incorporated and Cardillo teaches:

wherein processing each identified portion of the first set of audio signals comprises applying a computer- implemented speech recognition algorithm to data representing the first set of audio signals. (Page 11, Preprocessing, Acoustic Model, and Phonetic Grammar, ...*preprocessing engine scans the input speech and produces the corresponding phonetic search track*...)

As per claim 3, claim 1 is incorporated and Cardillo teaches:

wherein the subword units include linguistic units. (Page 11, Preprocessing, Acoustic Model, and Phonetic Grammar, *...preprocessing engine scans the input speech and produces the corresponding phonetic search track...*)

As per claim 4, claim 2 is incorporated and Cardillo fails to fully teach but Wolf teaches:

wherein locating the putative instances includes applying a computer-implemented word spotting algorithm configured using the representation of the spoken event of interest. (Cardillo teaches phonetic keyword spotting but fails to teach that the event of interest (query) is spoken. Wolf, ¶ 0054, *...a phoneme lattice, as described above, can also be used for devices with limited resources...In the case where the recognizer is part of the input device, e.g., a cell phone, the lattices can be forwarded to the search engine 190...*)

It would have been obvious to someone of ordinary skill in the art to combine Wolf with Cardillo to avoid losing information and adding ambiguity by application of a speech recognition process and performing traditional textual document retrieval. (Wolf, ¶ 0006)

As per claim 8, claim 1 is incorporated and Cardillo fails to fully teach but Wolf teaches:

wherein the representation of the spoken event of interest defines a network of subword units. (Wolf, ¶ 0054, *...a phoneme lattice, as described above, can also be used for devices with limited resources...In the case where the recognizer is*

part of the input device, e.g., a cell phone, the lattices can be forwarded to the search engine 190...)

It would have been obvious to someone of ordinary skill in the art to combine Wolf with Cardillo to avoid losing information and adding ambiguity by application of a speech recognition process and performing traditional textual document retrieval. (Wolf, ¶ 0006)

As per claim 9, claim 8 is incorporated and Cardillo fails to fully teach but Wolf teaches:

wherein the network of subword units is formed by multiple sequences of subword units that correspond to different paths through the network. (Wolf, ¶ 0054, *...a phoneme lattice, as described above, can also be used for devices with limited resources...In the case where the recognizer is part of the input device, e.g., a cell phone, the lattices can be forwarded to the search engine 190*...Figs 3a and 3b show the lattices.)

It would have been obvious to someone of ordinary skill in the art to combine Wolf with Cardillo to avoid losing information and adding ambiguity by application of a speech recognition process and performing traditional textual document retrieval. (Wolf, ¶ 0006)

As per claim 12, claim 1 is incorporated and Cardillo fails to fully teach but Wolf teaches:

accepting first audio data representing utterances of the event of interest spoken by a user, and processing the first audio data to form a processed query. (Wolf, ¶ 0054, *...a phoneme lattice, as described above, can also be used for devices with limited resources...In the case where the recognizer is part of the input device, e.g., a cell phone, the lattices can be forwarded to the search engine 190*...Figs 3a and 3b show the lattices.)

It would have been obvious to someone of ordinary skill in the art to combine Wolf with Cardillo to avoid losing information and adding ambiguity by application of a speech recognition process and performing traditional textual document retrieval. (Wolf, ¶ 0006)

As per claim 13, claim 1 is incorporated and Cardillo teaches:

accepting a selection by the user of portions of stored data from the first set of audio signals, and processing the portions of the stored data to form a processed query. (Cardillo, Page 14, teaches the use of the AudioLogger™ product for selecting media segments. Although the capabilities of AudioLogger™ are not discussed in Cardillo, NPL Document "Inventory of Metadata for Multimedia" teaches in section 3.4 that *...The Virage Tools automatically index your video content. The tools use speech recognition technology to extract information from the audio signal*... Therefore, the selection is an inherent feature of AudioLogger™ of which Cardillo incorporates by reference.)

As per claim 14, claim 13 is incorporated and Cardillo teaches:

prior to accepting the selection by the user, processing the first set of audio

signals according to a first computer-implemented speech recognition algorithm to produce the stored data. (Cardillo, Page 14, teaches the use of the AudioLogger™ product for selecting media segments. Although the capabilities of AudioLogger™ are not discussed in Cardillo, NPL Document "Inventory of Metadata for Multimedia" teaches in section 3.4 that ...*The Virage Tools automatically index your video content. The tools use speech recognition technology to extract information from the audio signal. A publishing application makes it possible to publish the searchable information*... Therefore, prior to the selection of the searchable information, a speech recognition algorithm is performed to produced the stored searchable data. The selection is an inherent feature of AudioLogger™ of which Cardillo incorporates by reference.)

As per claim 15, claim 14 is incorporated and Cardillo teaches:

wherein the first speech recognition algorithm produces data related to presence of the subword units at different times in the first set of audio signals. (Cardillo teaches in Fig. 1, a searching phase which locates putative instances of a query where ,Page 12, Temporal_Offset teaches a time location. Thus the collection of subword units generated by the first speech recognition algorithm is indicative of subword units at different times in the first set of audio signals.)

As per claim 16, claim 14 is incorporated and Cardillo teaches:

applying a second speech recognition algorithm to the processed query.

(Cardillo, Page 11, column 2, *...It is even conceivable that dynamic channel and language detection could be employed to switch acoustic models during preprocessing...* Furthermore, Page 12, column 1, *...A phonetic dictionary is probed for each word within the query term to accommodate unusual words (whose pronunciations must be handled specially for the given natural language...* It would have been obvious to someone of ordinary skill in the art at the time of the invention that if the audio is indexed in an alternative language and a person wants to search in a different language (North American English vs. Castilian Spanish), a second speech recognition algorithm with a phonetic dictionary specific to the language could have been applied to the processed query.)

Claims 17 and 18 are the computer readable medium and hardware representations of the method as claimed in claim 1. Claims 17 and 18 are rejected under the same principles as claim 1 for having similar limitations. Cardillo, Page 20, table 3 teaches computer implemented means for the method. The computer teaches a system, and the system cannot operate without being programmed, so a computer readable medium is inherent.

As per claim 19, claim 18 is incorporated and Cardillo fails to specifically teach, but Wolf teaches:

wherein the word spotter is further configured to identify time locations of the second audio signal at which the spoken event of interest is likely to have occurred

based on a comparison of the data representing the unknown speech with the representation of the spoken event of interest. (Cardillo teaches in Fig. 1, a searching phase which locates putative instances of a query where ,Page 12, Temporal_Offset teaches a time location. Cardillo fails to specifically teach a spoken event of interest (spoken query). Wolf, abstract, ...*A spoken query is represented as a lattice*...)

It would have been obvious to someone of ordinary skill in the art to combine Wolf with Cardillo to avoid losing information and adding ambiguity by application of a speech recognition process and performing traditional textual document retrieval. (Wolf, ¶ 0006)

8. Claims 5-7, 10-11 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cardillo et al. (NPL Document Phonetic Searching vs. LVCSR: How to Find What You Really Want in Audio Archives") in view of Wolf et al. (US PGPUB #20030204492) , and further in view of Ferrieux et al. (NPL Document "PHONEME-LEVEL INDEXING FOR FAST AND VOCABULARY-INDEPENDENT VOICE/VOICE RETRIEVAL")

As per claim 5, claim 4 is incorporated and Cardillo and Wolf fail to teach, but Ferrieux teaches:

selecting processing parameter values of the speech recognition algorithm for application to the data representing the first set of audio signals according to characteristics of the word spotting algorithm. (Section 2.2, ...*To compute the*

optimal parameters of those models,...)

It would have been obvious to someone of ordinary skill in the art at the time of the invention to combine Ferrieux with Cardillo and Wolf to optimize the parameters of the system based on the desired outcome (maximize likelihood/minimize error retrieval rate). (Ferrieux, section2.2)

As per claim 6, claim 5 is incorporated and Cardillo and Wolf fail to teach, but Ferrieux teaches:

wherein the selecting of the processing parameter values of the speech recognition algorithm includes optimizing said parameters according to an accuracy of the word spotting algorithm. (Section 2.2, ...*To compute the optimal parameters of those models,...maximize likelihood*...)

It would have been obvious to someone of ordinary skill in the art at the time of the invention to combine Ferrieux with Cardillo and Wolf to optimize the parameters of the system based on the desired outcome (maximize likelihood/minimize error retrieval rate). (Ferrieux, section2.2)

As per claim 7, claim 5 is incorporated and Cardillo and Wolf fail to teach, but Ferrieux teaches:

wherein the selecting of the processing parameter values of the speech recognition algorithm includes selecting values for parameters including one or more of an insertion factor, a recognition search beam width, a recognition grammar factor, and

a number of recognition hypotheses. (Ferriex, section 2.1 defines the models which include insertion costs, section 2.2 optimizes them.)

It would have been obvious to someone of ordinary skill in the art at the time of the invention to combine Ferrieux with Cardillo and Wolf to optimize the parameters of the system based on the desired outcome (maximize likelihood/minimize error retrieval rate). (Ferrieux, section2.2)

As per claim 10, claim 1 is incorporated and Cardillo and Wolf fail to teach, but Ferrieux suggests:

wherein forming the representation of the spoken event of interest includes determining an n-best list of recognition results. (Section 3, Synchronized Lattices, *...for each phoneme of the 1-best sequence, the posterior probabilities of the N-1 other phonemes on the same interval are also given...* Ferrieux further provides *...the single best sequence of phonemes often contains errors, and although systematic errors are handled by the confusion matrix, it is felt that knowledge of the second-best match would most often yield greater precision...* Therefore, given that the single best sequence of phonemes often contains errors, it would have been obvious to someone of ordinary skill in the art that multiple sequences through the lattice would have been beneficial to boost precision due to a higher number of sequences.)

It would have been obvious to someone of ordinary skill in the art at the time of the invention to combine Ferrieux with Cardillo and Wolf to optimize the parameters of the system based on the desired outcome (maximize likelihood/minimize error retrieval

rate). (Ferrieux, section2.2)

As per claim 11, claim 10 is incorporated and Cardillo and Wolf fail to teach, but Ferrieux suggests:

wherein each sequence of subword units in the representation corresponds to a different one in the n-best list of recognition results. (Section 3, Synchronized Lattices, *...for each phoneme of the 1-best sequence, the posterior probabilities of the N-1 other phonemes on the same interval are also given...* Ferrieux further provides *...the single best sequence of phonemes often contains errors, and although systematic errors are handled by the confusion matrix, it is felt that knowledge of the second-best match would most often yield greater precision...* Therefore, given that the single best sequence of phonemes often contains errors, it would have been obvious to someone of ordinary skill in the art that multiple sequences through the lattice would have been beneficial to boost precision due to a higher number of sequences. The phoneme sequences are the matches (which defines the first and second options) in Ferrieux, therefore it would have further been obvious that the phoneme sequences through the lattice would have been each of the different options in the n-best criterion evaluations.)

It would have been obvious to someone of ordinary skill in the art at the time of the invention to combine Ferrieux with Cardillo and Wolf to optimize the parameters of the system based on the desired outcome (maximize likelihood/minimize error retrieval rate). (Ferrieux, section2.2)

Conclusion

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Refer to PTO-892, Notice of References Cited for a listing of analogous art.

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to GREG A. BORSETTI whose telephone number is (571)270-3885, (FAX: 571-270-4885). The examiner can normally be reached on Monday - Thursday (8am - 5pm Eastern Time).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, RICHEMOND DORVIL can be reached on 571-272-7602. The fax phone

number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Greg A. Borsetti/
Examiner, Art Unit 2626

/Talivaldis Ivars Smits/
Primary Examiner, Art Unit 2626 6/22/2010